Martens et al. Supplementary data

Supplementary Methods

Cell culture

Kasumi-1 (Asou et al., 1991), SKNO-1 (Matozaki et al., 1995) and U937 AML1-ETO (UAE) cells (Alcalay et al., 2003) were routinely cultured in RPMI 1640 supplemented with FCS at 37 °C. SKNO-1 cells were cultured in the presence of GM-CSF (Sigma). AML1-ETO expression in UAE cells was induced by treatment for 5 hours with 1 mM zinc. 293T and MCF7 cells were cultured in DMEM supplemented with 10% FCS at 37 °C. K562-ERG (inducible expressing ERG3) cells were cultured in RPMI with 10% FCS, 500 μg/ml G418 and 1 μg/ml puromycin at 37 °C. ERG expression in K562-ERG cells was induced by treatment for 72 hours with 1 μg/ml doxycyclin.

Antibodies

ETOsc sc-9737, Santa Cruz
HEB sc-357, Santa Cruz
ERG sc-353, Santa Cruz
ERG sc-354, Santa Cruz (used in ChIP-seq in Yu et al., 2010; Wilson et al., 2010)
FLI1 sc-356, Santa Cruz (used in ChIP-seq in Wei et al., 2010)
H3K9K14ac, pAb-ACHBHS-044_DA-0010, Diagenode
AML1-ETO, A706, Diagenode
ETO1, A710, Diagenode
CBFβ, A1329, Diagenode
RNAPII, 8WG16, Diagenode
TBP, MAb-002-100, Diagenode
RUNX1, ab-23980, Abcam (used in ChIP-seq in Tijssen et al., 2011, Wilson et al., 2010)
KAP1, ab-10484, Abcam
FLI1, ab-15289, Abcam (used in ChIP-seq in Tijssen et al., 2011; Wilson et al., 2010)
RUNX1, PC285, Calbiochem
Actin, sc-1646, Santa Cruz

Chromatin immunoprecipitation (ChIP)

Chromatin was harvested as described (Denissov et al., 2007). ChIPs were performed using specific antibodies to ETO, HEB, ERG, FLI1 (Santa Cruz), H3K9K14ac, AML1-ETO, ETO, CBFβ, RNAPII (Diagenode), RUNX1, FLI1 (Abcam) and H4panAc (Millipore) and analyzed by quantitative PCR (qPCR) or ChIP-seq. Primers for qPCR are described below. Relative occupancy was calculated as fold over background, for which the second exon of the *Myoglobin* gene or the promoter of the *H2B* gene was used. For sequential ChIP, chromatin of the first ChIP was eluted from the beads in 100 μl elution buffer (1% SDS), diluted 10 times in incubation buffer (50 mM Tris pH 8.0; 100 mM NaCl; 2 mM EDTA; 1 mM DTT; 1% NP40; protease inhibitor cocktail, Roche) and a second round of ChIP was performed.

Transfection
293T, MCF7 and K562-ERG cells were transfected with pcDNA-ERG1 or -AML1-ETO expression constructs using lipofectamine (Invitrogen) according to the manufacturers protocol. Cells were harvested 24 hours after transfection. Protein lysates were tested by western blotting using antibodies against AML1-ETO (AE), TBP (Diagenode), KAP1 (Abcam) or ERG (sc-353, Santa Cruz) and subsequently used for ChIP experiments**.**

Co-immunoprecipitation
Coimmunoprecipitation experiments were performed as before (Martens et al., 2002) in assay buffer (0.1% NP-40, 250 mM NaCl, 50 mM Tris-HCl (pH 7.5) containing a mixture of protease inhibitors). SKNO-1 protein lysates were incubated overnight with ERG or IgG antibodies and prot A/G beads (Santa Cruz), washed 4 times in assay buffer and tested using western blotting for the presence of AML1-ETO or RNAPII (Diagenode).

GST-fusion proteins
GST fusion protein-coated beads and GST fusion proteins were prepared as previously reported (Martens et al., 2002). GST fusion proteins were constructed by PCR amplification of different AML1-ETO domains in pGEX-2T using the BamHI and EcoRI restriction sites. Expression of GST and GST-fusion proteins was induced by IPTG treatment for 3 hours.
GST-constructs (with corresponding AML1-ETO amino acid sequence):
1 RHD/AML (aa 1-183)
2 PST1 (aa 172-271)
3 NHR1 (aa 257-395)
4 PST2 (aa 396-481)
5 NHR2 (aa 467-579)
6 NHR3 (aa 565-662)
7 NHR4PST (aa 663-752)

MethylCap™
Pull down experiments were performed using GST fused to the MBD domain of MeCP2 (Diagenode). DNA was isolated from blast cells, sonicated to generate fragments of approximately 400 bp and pulled down with GST-MBD coated paramagnetic beads and the IP-STAR robot (Diagenode). After washing with 200 mM NaCl, the bound methylated DNA was eluted using 700 mM NaCl and used for high-throughput DNA sequencing (Brinkman et al., 2010).

Bioinformatic analysis

Identification of AML1-ETO binding sites in Kasumi-1 and SKNO-1
AML1-ETO peaks in SKNO-1 and Kasumi-1 cells were detected using MACS (Zhang et al., 2008) at a p-value cut off for peak detection of 10^{-8} (Supplementary Tables, page 1). To identify high confidence binding sites, i.e., the strongest fraction of binding events in both these cell lines we employed a regression analysis in which each binding site is evaluated for its relative tag density in both cell lines. For this, in each resulting peak

region the number of tags for AML1-ETO in Kasumi-1 and SKNO-1 cells was counted. Subsequently all regions were tested for relative AML1-ETO tag densities (tag density at peak divided by total number of tags in all peaks), sorted and visualized in a dot plot. The data points of the dot plot were subsequently used for regression analysis, with resulting regression curves, plus cut off values shown in figure 1B. To increase visibility, dots representing the individual data points were removed. A cut off value was set at 0,00010 (>14 tags/kb), which represent in Kasumi-1 cells a binding site composed of 14 tags in a window of 1 kb and 6.2 million tags sequenced in total.

Quantitative PCR validation of AML1-ETO binding sites
High confidence AML1-ETO peaks from Kasumi-1/SKNO-1 cells were divided in three categories: high, middle, low. From each of these categories 10 peaks were selected and subsequently validated in ChIP-qPCR experiments using the primer pairs below. The resulting occupancy levels for each of the three categories was plotted in a boxplot and compared to the number of tags within each high, middle or low peak region.

Peak detection
Peaks were generally identified using MACS (Zhang et al., 2008). Random genomic regions were selected using the complete human genome sequence and the Rand function of Perl to identify sets of random genomic positions. These random positions were subsequently extended to 1 kb.

Tag counting
Tags within a given region were counted and adjusted to present the number of tags within a 1 kb region. Subsequently the percentage of these tags as a measure of the total number of sequenced tags of the sample was calculated. For the heatmap display in Figures 1C and S2B a cut off was used of 3 % tags/kb (10^{-4}), which represent a peak of 1000 bp width and composed of 30 tags or more with 10 million tags sequenced (or 15 tags with 5 million tags sequenced). In Figure 7H and I the average tag density per bin of H3ac or DNAme from two patients (pz229 and pz186) was determined. In Supplementary Figure S4C a t-test was used to show statistical difference between AML1-ETO occupancy before and after dox treatment of K562-ERG cells. For box plots the middle dot represents the median value, the bottom and top of the box are the 25^{th} and 75^{th} percentile and the ends of the whiskers represent the 9^{th} and the 91^{st} percentile.

Peak distribution analysis
To determine genomic locations of binding sites peak files were analyzed using a script that annotates binding sites according to all RefSeq genes. With this tool every binding site is annotated either as promoter (-500 bp to the Transcription Start Site), non promoter CpG island, intron, exon or intergenic (everything else).

Accessibility mapping
To examine whether ERG binds to accessible sites we used public available DNAseI accessibility data from K562 cells (GEO series GSE29692) and the DNAseI hotspots as can be found under the ‘regulation’ tracks in the UCSC browser.

Motif analysis

To identify the motifs underlying the AML1-ETO peaks gimmemotifs (van Heeringen and Veenstra, 2011) was used. Briefly, gimmemotifs is a *de novo* motif prediction pipeline combining three motif prediction tools, MotifSampler (Thijs et al., 2001), Weeder (Pavesi et al., 2004) and MDmodule (Liu et al., 2002). Gimmemotifs was run on 20% of randomly selected 200-bp peak sequences (centered at the peak summit as reported by MACS) and position weight matrices (PWMs) were generated. The 'large' analysis setting was used for Weeder. MDmodule and MotifSampler were each used to predict 10 motifs for each of the widths between 6 and 20. The significance of the predicted motifs was determined by scanning the remaining 80% of the peak sequences and two different backgrounds: a set of random genomic sequences with a similar genomic distribution as the peak sequences and a set of random sequences generated according to a 1st order Markov model, matching the dinucleotide frequency of the peak sequences. P-values were calculated using the hypergeometric distribution with the Benjamin-Hochberg multiple testing correction. All motifs with a p-value <0.001 and an absolute enrichment of at least >1.5-fold compared to both backgrounds were determined as significant.

To count motifs in ERG binding sites we derived the weight matrix of different consensus binding sites for various proteins involved in hematopoiesis from Jaspar (http://jaspar.genereg.net/). All ERG binding sites were subsequently examined for the presence or absence of these motifs using a script that scans for homology of the matrix within the DNA sequence underlying the ERG binding site (pwmscan.py)(see also van Heeringen et al., 2010) using a threshold score of 0.9 (on a scale from 0 to 1). Due to the different composition and length of the motifs the resulting homology scores could not be directly compared and needed to be normalized. For this we calculated the lower (no homology) and higher (complete homology) scores for each individual motif with the script pwm_scores.py (available upon request), and used these scores to rank each motif score within an ERG binding site to a scale from 0 to 1. These ranked values were subsequently displayed in a heatmap in which red means a high score for a particular motif (and thus the presence of a motif) and green a low score (and the absence of the motif).

For the motif count distribution analysis in Figure 6F the Chi-square test was used to show a statistical significant change in the pattern.

Identification of AML1-ETO binding sites in patient cells

Peaks in patient samples 12, 186 and 229 were detected using MACS (Zhang et al., 2008) at a p-value cutoff for peak detection of 10^{-6}. Resulting peak files were overlapped and common peaks identified in all three patient samples were selected for further analysis.

Expression analysis

Total RNA was extracted from SKNO-1 cells with the RNeasy kit and on-column DNase treatment (Qiagen) and the concentration was measured with a Qubit fluorometer (Invitrogen). 250 ng of total RNA was treated by Ribo-Zero rRNA Removal Kit (epicentre) to remove ribosomal RNAs according to manufacturer instructions. 16 µl of purified RNA was fragmented by addition of 4 µl 5x fragmentation buffer (200 mM Tris acetate pH 8.2, 500 mM potassium acetate and 150 mM magnesium acetate) and

incubated at 94°C for exactly 90 seconds. After ethanol precipitation first strand cDNA was synthesized from the fragmented RNA with SuperscriptIII (Invitrogen) using random hexamers. First strand cDNA was purified by Qiagen mini elute columns and second strand cDNA was prepared in the presence of dUTP instead of dTTP. Double stranded cDNA was purified by Qiagen mini elute columns and used for Illumina sample prepping and sequenced according to the manufacturer's instructions. A total of 16,178,852 RNA-seq reads were uniquely mapped to HG18 and used for bioinformatic analysis. RPKM (reads per kilobase of gene length per million reads) (Mortazavi et al., 2008) values for RefSeq genes were computed using tag counting scripts and used to analyze the expression level of genes in SKNO-1 cells. A t-test was used to show statistical difference between expression of groups of genes.

shRNA-mediated silencing of ERG in SKNO-1 cells was performed using two shRNA with the following target sequences: GACTCTTGGGAGGGAGTTA (shRNA-ERG1) and CGACATCCTTCTCTCACAT (shRNA-ERG2) (Tsuzuki et al., 2011). Briefly shRNA was cloned in Tet-pLKO-puro vector and transfected in HEK293T cells to produce virus. Viral supernatants were harvested at 48h and 72 h for transduction of SKNO-1 cells pretreated with 6ug/ml polybrene. Cells were subsequently selected using 2 ug of puromycin for 2 weeks. shRNA expression was induced by 1 ug/ml doxycycline for 72 hours.

Expression of ETS factors in AML samples was examined in the dataset published by Valk et al., (2004) using oncomine (www.oncomine.com). ETS factor candidate proteins were selected based on levels of expression and change in expression in AML as compared to control cells ($CD34^+$ and bone marrow).

Expression of ERG and FLI1 in SKNO-1, UAE and NB4 cells was examined using RNA-seq (data not shown) and analyzing RPKM values. This revealed that SKNO-1 cells express both ERG and FLI1 to equal level, while in UAE and NB4 cells FLI1 is highly expressed while ERG is not detectable (RPKM value in both cells is 0).

t(8;21) patient characteristics

t(8;21) AML blasts from peripheral blood or bone marrow from *de novo* AML patients were studied after informed consent was obtained in accordance with the Declaration of Helsinki. None of the three patients (12, 186 and 229) had a FLT3-ITD mutation, one patient (12) had a c-Kit mutation.

Primers used in this study

ChIP:

SPI1	Forward	GGGTAAGAGCCTGTGTCAGC
	Reverse	CAGATGCACGTCCTCGATAC
FUT7	Forward	TGAAACCAACCCTCAAGGTC
	Reverse	TCACTGGCATGAATGAGAGC
NFE2	Forward	GGTTAGCAGCATACGTGGAG
	Reverse	ACGATACGGAGAAAACCACG
OGG1	Forward	CCACCCTGATTTCTCATTGG

	Reverse	CAACCACCGCTCATTTCAC
VEGF	Forward	GGTTTGGATCCTCCCATTTC
	Reverse	CAGTCAGTGGTGGGGAGAG
ITGAM	Forward	GCTTCCTTGTGGTTCCTCAG
	Reverse	AGGAGCCAGAACCTGGAAG
CD344	Forward	AGTTTGGCTTGTGGGAACTG
	Reverse	GACAAGGCCACTGAGAAAGC
KREMEN1	Forward	CGAGAGTGACATCCAGTTGC
	Reverse	TTCACAACCGTTCCAGATGA
H2B	Forward	TTGCATAAGCGATTCTATATAAAAGCG
	Reverse	ATAAAGCGCCAACGAAAAGG
MYOG	Forward	AAGTTTGACAAGTTCAAGCACCTG
	Reverse	TGGCACCATGCTTCTTTAAGTC

Cloning GST fusion proteins AML1-ETO:

GST-1	forward	ACTGCGGATCCCGTATCCCCGTAG
	reverse	CAGTGAATTCTCAGTGCTTCTCAG
GST-2	forward	ACTGCGGATCCGGGCCCCGAGAACCTC
	reverse	CAGTGAATTCTCAGAGTTGCCTGGC
GST-3	forward	ACTGCGGATCCCTGGCTAATCAACAG
	reverse	CAGTGAATTCTCATCTGTCTGGAGTTC
GST-4	forward	ACTGCGGATCCACCAAAGAAAATGGC
	reverse	CAGTGAATTCTCAATGCAACCCCATAG
GST-5	forward	ACTGCGGATCCAGCCACAGGGAC
	reverse	CAGTGAATTCTCATTCCCGATGCGC
GST-6	forward	ACTGCGGATCCAGTCCCGTCAACC
	reverse	CAGTGAATTCTCAACTCTCGCTTGAATC
GST-7	forward	ACTGCGGATCCTGCTGGAATTGTG
	reverse	CAGTGAATTCTCACTAGCGAGGGGTTG

qPCR validation of high confidence AML1-ETO SKNO-1/Kasumi-1 binding sites

High regions		
1	forward	AAGGGAGGGGAGCTAACTGA
	reverse	GGCTAATCCCACAGAGCAAG
2	forward	CACGCTGGCTACATTTCTCA
	reverse	GTGTCCCCTCTTGCTGACAT
3	forward	CTTCAGTGGCAAACCCAGTT
	reverse	GCAAGGAAGCTGAGGATGAG
4	forward	TGTGTTGGTTGGAAGCTGAA
	reverse	AAGACCTGTTGCCAGCATCT
5	forward	TTGTTGGGGAACACTTCACA
	reverse	AAGGCTGAGAAAAGGGAAGC
6	forward	CTGGACTGGGGAAGGATTTT

	reverse	ACCCCACACACACTCCCTTA
7	forward	AAATGGCAACTGGACCAAAG
	reverse	GTCGACATCTCCTCCAGCTC
8	forward	TCCACAGAAGCCTCCTTGTT
	reverse	TTGTTTCACCACCAGACTGC
9	forward	AATTGCTGTGCACTGTGTCC
	reverse	GACCACAGCATCCCATTCTT
10	forward	CCAAGTTTGCGCAATAGGAC
	reverse	CCATGTGCCTTGCACAATAA
Middle regions		
1	forward	GGCCACACTTCATTTCACCT
	reverse	TAGCGGGAGAGGCAGAGATA
2	forward	TGACGCTTAAGAGCCCAGAT
	reverse	AGCAAGACCACTGCTGGAAT
3	forward	CAGCTTGTTTGCACTTTGGA
	reverse	AGCAGCCTGACTTGAAAAGC
4	forward	GGGTCACATCTCCTCCTTCA
	reverse	GCCACTCAAGCTCACTCTCC
5	forward	GCATTTGGAGGCTACTGCTC
	reverse	TCGGAGGTGAGAATGCTCTT
6	forward	TCTGCTGACAACCTGAATGC
	reverse	GGCTTAGGATGGGGGAGTAG
7	forward	AGAGCTCAGGTGTCGTCCAT
	reverse	GCAAACTGAGCTGTGGCATA
8	forward	ACAGGCATCTCCCAGCTCTA
	reverse	CTTGTGTGCTGGAGGTTGTG
9	forward	TCTCCAAGCAGCTGATGATG
	reverse	AGATGAATGGGAGGGAGCTT
10	forward	GGGAAAGGTCCAGAGAGAGG
	reverse	TGTCTGGAAGGGGAATTCAG
Low regions		
1	forward	GCTGGCAGTTAAGGGATGAG
	reverse	CTCTAGCTGCTGCCCTGTCT
2	forward	AAGCTGGAGAACAAGGCTCA
	reverse	GTCAGGGGGTGACACAGACT
3	forward	TCCTACGTTCTGCCCATTGT
	reverse	CTCCCAAAGAGTTGCCAGAC
4	forward	GAGAGACTGCTGCGGGTAAC
	reverse	GCTTCTGCAAAGCCTGACTC
5	forward	CACCAGCCTGAACAGATGAA
	reverse	TCCAAACAGCAAAGGAGCTT
6	forward	GAATCTGGGTGTTGCAAGGT
	reverse	GGTGATCCTAGGGGGAGAAG
7	forward	CTGGGACGTGAAGAGGAGAC
	reverse	AGAGCCTTACAATGCCTGGA
8	forward	TTCCTATGGACTCCCACAGC
	reverse	AGTCCATGGGGCAGTAGATG
9	forward	GGACTTCCAGGCCATGACTA
	reverse	TCCTTCTCTTTGGGGTCCTT

10	forward	GCAGAGCTTGTGGGAGTTTC
	reverse	CAGAGAGACACGCCTGTACG

Profiles analyzed in this study

Cells	ChIP antibody/technique	Treatment	Mapped reads	reference
Kasumi-1	AE (A706)	no	6716821	
Kasumi-1	HEB (sc-357)	no	5885202	
Kasumi-1	ETO1 (A710)	no	6738375	
Kasumi-1	ETOsc (sc-9737)	no	5193085	
Kasumi-1	H3K9ac	siAML1-ETO		Ptasinska et al., 2012
Kasumi-1	H3K9ac	siControl		Ptasinska et al., 2012
SKNO-1	AE (A706)	no	9474494	
SKNO-1	CBFβ (A1329)	no	2084211	
SKNO-□	ERG (sc-353)	no	13373986	
SKNO-1	FLI1 (sc-356)	no	16093275	
SKNO-1	RUNX1 (ab-23980)	no	2084211	
SKNO-1	RNA-seq		16178852	
SKNO-1	RNA-seq shERG1		56302392	
SKNO-1	RNA-seq shERG2		20711638	
AML pz12	AE (A706)	no	11324391	
AML pz12	ERG (sc-353)	no	16659875	
AML pz186	AE (A706)	no	8342982	
AML pz186	H3K9K14ac (Diagenode)	no	10175724	
AML pz186	MethylCap	no	34716102	
AML pz229	AE (A706)	no	8882375	
AML pz229	H3K9K14ac (Diagenode)	no	15944616	
AML pz229	MethylCap	no	21305015	
CD34+ nr29	ERG (sc-353)	no	16965117	
CD34+ nr30	H3K9K14ac (Diagenode)	no	16201598	
CD34+ nr30	FLI1 (sc-356)	no	16191803	
NB4	PML (H238)	no		Martens et al., 2010
NB4	RARa (Diagenode)	no		Martens et al., 2010
NB4	FLI1 (ab-15289)	no	8935568	
APL pz74	ERG (sc-353)	no	17758130	
MCF7	ERG (sc-353)	AML1-ETO/ERG transfected	3544120	
MCF7	AE (A706)	AML1-ETO/ERG transfected	3648588	
K562-ERG	ERG (sc-353)	AML1-ETO transfected, no dox	23220874	
K562-ERG	ERG (sc-353)	AML1-ETO transfected,72 hrs dox	18572309	
K562-ERG	AE (A706)	AML1-ETO transfected, no dox	13700478	
K562-ERG	AE (A706)	AML1-ETO	12259664	

		transfected,72 hrs dox		
UAE	AE (A706)	no	8277859	
UAE	AE (A706)	5 hrs zinc	7670219	
UAE	FLI1 (ab-15289)	no	17661457	
UAE	FLI1 (ab-15289)	5 hrs zinc	19010064	
UAE	H3K9K14ac (Diagenode)	no	14723129	
UAE	H3K9K14ac (Diagenode)	5 hrs zinc	15057351	
UAE	H4panac (Upstate)	no	13195517	
UAE	H4panac (Upstate)	5 hrs zinc	11996351	
CD133+	H3K4me3	no		Cui et al., 2009
CD133+	H3K9me1	no		Cui et al., 2009
CD133+	H3K9me3	no		Cui et al., 2009
CD133+	H3K27me1	no		Cui et al., 2009
CD133+	H3K27me3	no		Cui et al., 2009
CD133+	H4K20me1	no		Cui et al., 2009
CD133+	H3K4me1	no		Cui et al., 2009
CD133+	H3K36me3	no		Cui et al., 2009

Scripts used in this study

Task	Name script	Used to generate figures
Peak calling	MACS	1B, H; 2E; 3A; 5B, D; 6D, E; S1H, I; S3C; S7B
Tag counting	peakstats.py	1B, C, G; 2E; 3E; 7A, C-I; S1E, G; S2B; S3C, S4C
Motif discovery	gimme_motifs.py	1E
Motif counting	pwmscan.py	6C,F; S7A
Motif scoring	pwm_scores.py	6C; S7A
Peak annotation	genomic_distribution.sh	1D, 3D
Intensity plot	makeColorProfiles.pl	1G; 2E, 7D

All scripts used in this study are available upon request

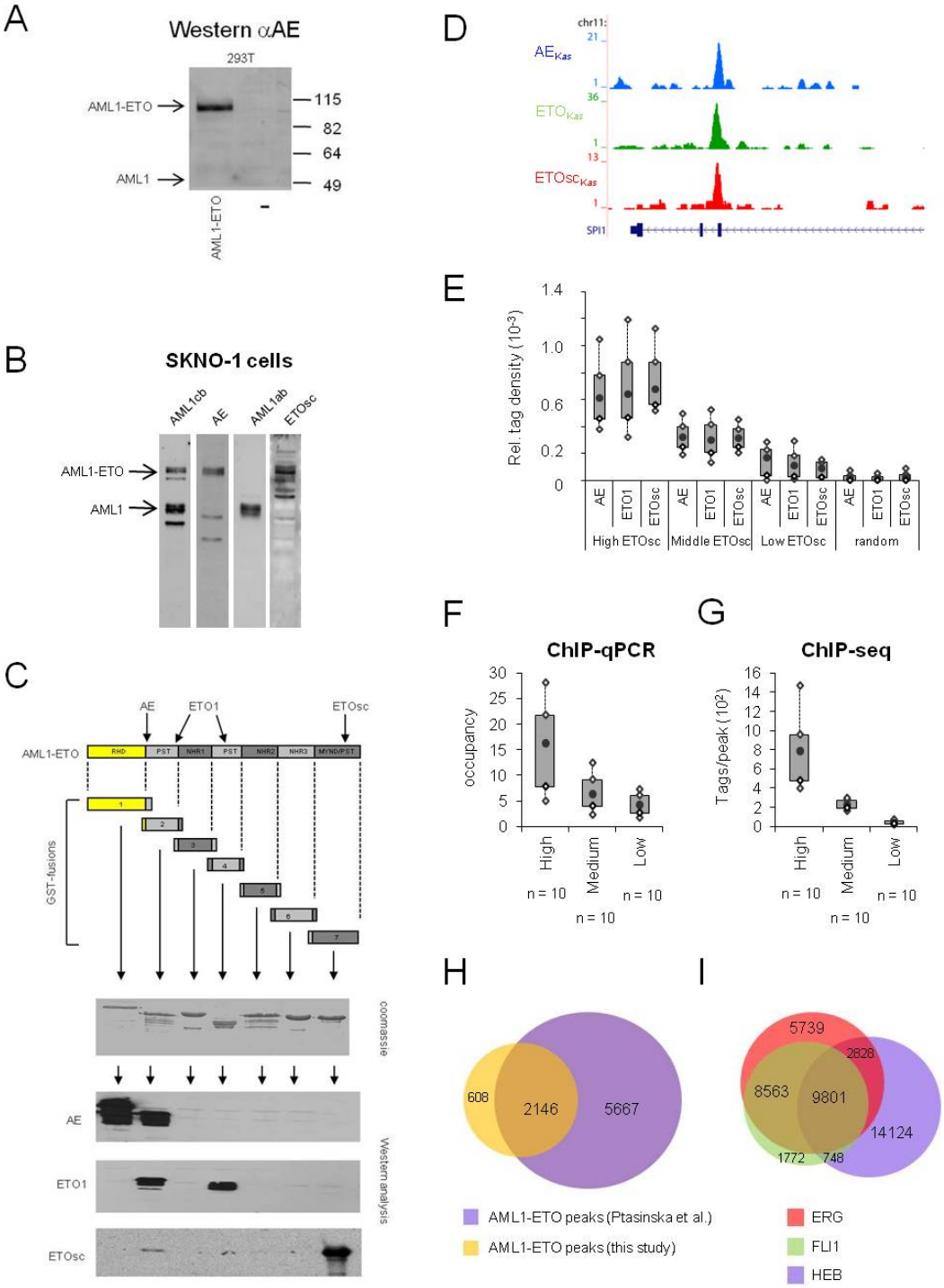
For clustering and heatmap generation TMEV (http://www.tm4.org/mev/) was used and for functional annotation DAVID (http://david.abcc.ncifcrf.gov/).

Supplementary References

1. Alcalay M, Meani N, Gelmetti V, Fantozzi A, Fagioli M, Orleth A, *et al.* Acute myeloid leukemia fusion proteins deregulate genes involved in stem cell maintenance and DNA repair. *J Clin Invest* 2003 Dec; **112**(11)**:** 1751-1761.
2. Asou H, Tashiro S, Hamamoto K, Otsuji A, Kita K, Kamada N. Establishment of a human acute myeloid leukemia cell line (Kasumi-1) with 8;21 chromosome translocation. *Blood* 1991 May 1; **77**(9)**:** 2031-2036.
3. Brinkman AB, Simmer F, Ma K, Kaan A, Zhu J, Stunnenberg HG. Whole-genome DNA methylation profiling using MethylCap-seq. *Methods* 2010 Jun 10.
4. Cui K, Zang C, Roh TY, Schones DE, Childs RW, Peng W, *et al.* Chromatin signatures in multipotent human hematopoietic stem cells indicate the fate of bivalent genes during differentiation. *Cell Stem Cell* 2009 Jan 9; **4**(1)**:** 80-93.
5. Denissov S, van Driel M, Voit R, Hekkelman M, Hulsen T, Hernandez N, *et al.* Identification of novel functional TBP-binding sites and general factor repertoires. *Embo J* 2007 Feb 21; **26**(4)**:** 944-954.
6. Liu XS, Brutlag DL, Liu JS. An algorithm for finding protein-DNA binding sites with applications to chromatin-immunoprecipitation microarray experiments. *Nat Biotechnol* 2002 Aug; **20**(8)**:** 835-839.
7. Martens JH, Brinkman AB, Simmer F, Francoijs KJ, Nebbioso A, Ferrara F, *et al.* PML-RARalpha/RXR Alters the Epigenetic Landscape in Acute Promyelocytic Leukemia. *Cancer Cell* 2010 Feb 17; **17**(2)**:** 173-185.
8. Martens JH, Verlaan M, Kalkhoven E, Dorsman JC, Zantema A. Scaffold/matrix attachment region elements interact with a p300-scaffold attachment factor A complex and are bound by acetylated nucleosomes. *Mol Cell Biol* 2002 Apr; **22**(8)**:** 2598-2606.
9. Matozaki S, Nakagawa T, Kawaguchi R, Aozaki R, Tsutsumi M, Murayama T, *et al.* Establishment of a myeloid leukaemic cell line (SKNO-1) from a patient with t(8;21) who acquired monosomy 17 during disease progression. *Br J Haematol* 1995 Apr; **89**(4)**:** 805-811.

10. Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat Methods* 2008 Jul; **5**(7)**:** 621-628.
11. Pavesi G, Mereghetti P, Mauri G, Pesole G. Weeder Web: discovery of transcription factor binding sites in a set of sequences from co-regulated genes. *Nucleic Acids Res* 2004 Jul 1; **32**(Web Server issue)**:** W199-203.
12. Ptasinska A, Assi SA, Mannari D, James SR, Williamson D, Dunne J*, et al.* Depletion of RUNX1/ETO in t(8;21) AML cells leads to genome-wide changes in chromatin structure and transcription factor binding. *Leukemia* 2012 Feb 20.
13. Thijs G, Lescot M, Marchal K, Rombauts S, De Moor B, Rouze P*, et al.* A higher-order background model improves the detection of promoter regulatory elements by Gibbs sampling. *Bioinformatics* 2001 Dec; **17**(12)**:** 1113-1122.
14. Tijssen MR, Cvejic A, Joshi A, Hannah RL, Ferreira R, Forrai A*, et al.* Genome-wide analysis of simultaneous GATA1/2, RUNX1, FLI1, and SCL binding in megakaryocytes identifies hematopoietic regulators. *Dev Cell* 2010 May 17; **20**(5)**:** 597-609.
15. Tsuzuki S, Taguchi O, Seto M. Promotion and maintenance of leukemia by ERG. *Blood* 2011 Apr 7; **117**(14)**:** 3858-3868.
16. Valk PJ, Verhaak RG, Beijen MA, Erpelinck CA, Barjesteh van Waalwijk van Doorn-Khosrovani S, Boer JM*, et al.* Prognostically useful gene-expression profiles in acute myeloid leukemia. *N Engl J Med* 2004 Apr 15; **350**(16)**:** 1617-1628.
17. van Heeringen SJ, Veenstra GJ. GimmeMotifs: a de novo motif prediction pipeline for ChIP-sequencing experiments. *Bioinformatics* 2011 Jan 15; **27**(2)**:** 270-271.
18. Wang L, Gural A, Sun XJ, Zhao X, Perna F, Huang G*, et al.* The leukemogenicity of AML1-ETO is dependent on site-specific lysine acetylation. *Science* 2011 Aug 5; **333**(6043)**:** 765-769.
19. Wei GH, Badis G, Berger MF, Kivioja T, Palin K, Enge M*, et al.* Genome-wide analysis of ETS-family DNA-binding in vitro and in vivo. *EMBO J* 2010 Jul 7; **29**(13)**:** 2147-2160.
20. Wilson NK, Foster SD, Wang X, Knezevic K, Schutte J, Kaimakis P*, et al.* Combinatorial transcriptional control in blood stem/progenitor cells: genome-wide analysis of ten major transcriptional regulators. *Cell Stem Cell* 2010 Oct 8; **7**(4)**:** 532-544.
21. Yu J, Mani RS, Cao Q, Brenner CJ, Cao X, Wang X*, et al.* An integrated network of androgen receptor, polycomb, and TMPRSS2-ERG gene fusions in prostate cancer progression. *Cancer Cell* 2010 May 18; **17**(5)**:** 443-454.
22. Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE*, et al.* Model-based analysis of ChIP-Seq (MACS). *Genome Biol* 2008; **9**(9)**:** R137.

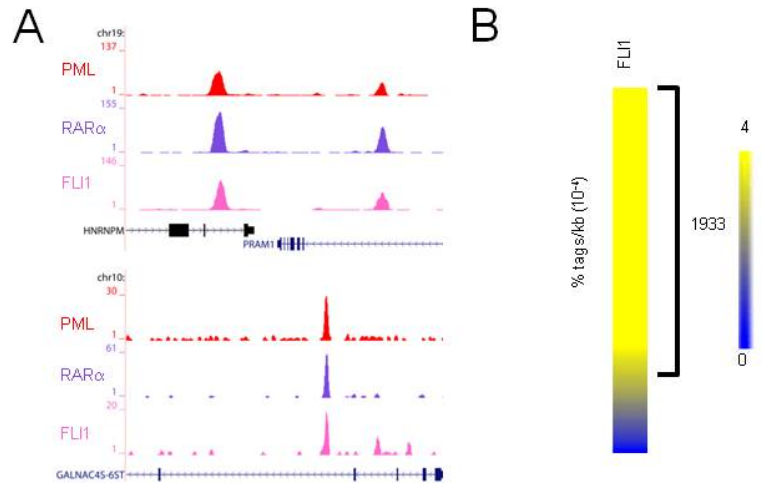
Figure S1



Supplementary Figure S1. A. 293T cells were transfected with full length AML1-ETO or control vector and protein extracts were analyzed for the presence of AML1-ETO in Western using the AE antibody. Only

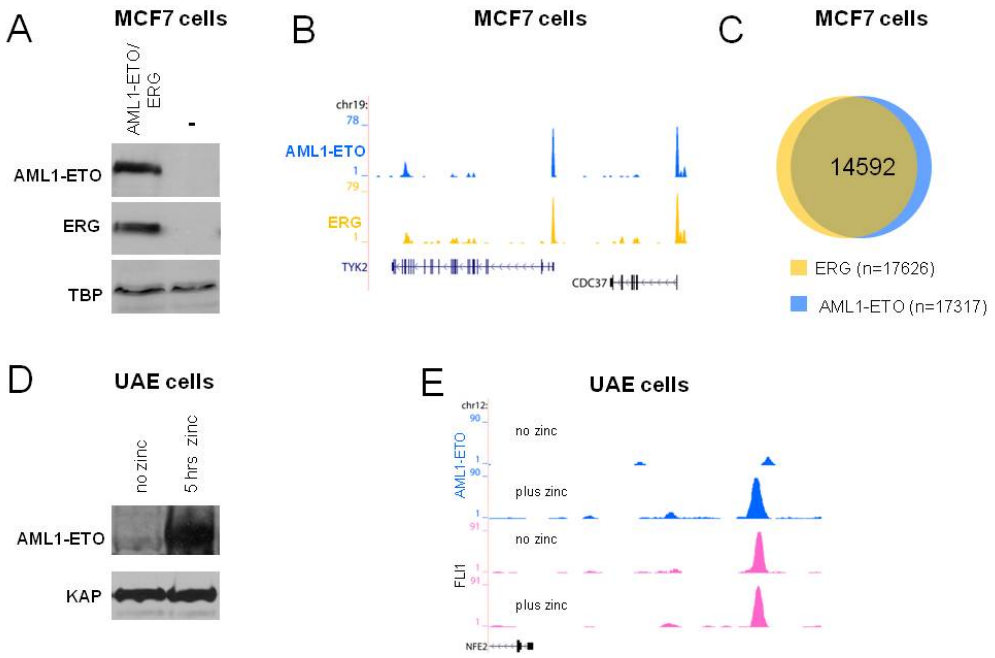
a signal was detected at the expected height of the AML1-ETO protein. B. Western blot analysis comparing the recognition capacity of an N-terminal RUNX1 antibody that recognizes both the wt as well as the fused RUNX1 (AMLcb; Calbiochem PC285), the AML1-ETO antibody used in this study (AE, A706), the RUNX1/AML1 antibody used for ChIP-seq in this study that recognizes the middle domain of RUNX1 which is not retained in the AML1-ETO fusion protein (AML1ab, Abcam ab-23980) and the ETO C-terminal antibody (ETOsc) used for ChIP-seq in Ptasinska et al. (2012) and Wang et al. (2011). C. Analysis of the recognition capacity of the AE, ETO1 (Diagenode) and ETOsc (Santa Cruz) antibodies towards GST fusion domains of AML1-ETO. The fusion point of AML1-ETO is present both in GST fusion product 1 and 2, peptides that were used for generating the ETO1 antibody were present in GST fusion 2 and 4 while the Santa Cruz ETO antibody (ETOsc) was developed against a peptide present in GST fusion 7. RHD, Runt-Homology Domain; PST, Proline-Serine-Threonine-rich region; NHR, Nervy Homology Region; MYND, Myeloid-Nervy-Deaf domain. The RUNX1/AML1 part of AML1-ETO is highlighted in yellow. D. Analysis of AML1-ETO binding sites using two additional ETO antibodies, ETO1 and ETOsc. Overview of the *SPI1* AML1-ETO binding site in Kasumi-1 cells. In blue the AE ChIP-seq data is plotted, in green the ETO1 and in red the ETOsc data. E. Boxplot showing the percentage of AE, ETO1 and ETOsc tags, within three subgroups (of the 2,754 AML1-ETO binding sites) that harbor different ETOsc densities or a set of random regions of similar size. F. Validation of ChIP-sequencing data by qPCR. Randomly high, medium and low (n=10) AML1-ETO binding sites were selected and validated for AML1-ETO binding by ChIP-qPCR in SKNO-1 cells. Occupancy results for each class of binding sites (high, medium, low) are represented in a boxplot. G. SKNO-1 AML1-ETO ChIP-seq tag count for the selected high, medium and low binding sites. H. Venn diagram showing an overlap of 80% of the high confidence AML1-ETO binding sites identified using the combination of SKNO-1 and Kasumi-1 cells as presented in this study with those identified in Kasumi-1 cells in Ptasinska et al. (2012). I. Venn diagram representing the overlap of ERG, FLI1 and HEB binding sites.

Figure S2



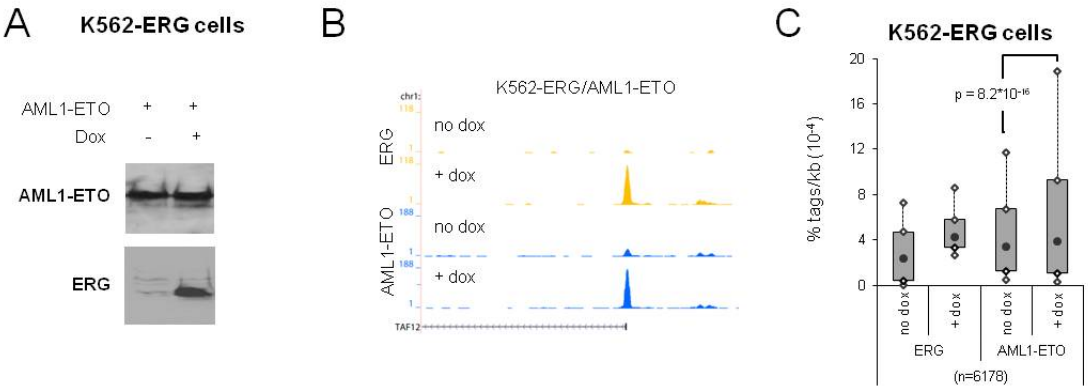
Supplementary Figure S2. A. Overview of the *PRAM1* and *GALNAC4S-6ST* genes in NB4 cells. In red the PML, in purple the RARα and in pink the FLI1 ChIP-seq data is plotted. B. Heat map displaying FLI1 tag densities at high confidence PML-RARα binding sites.

Figure S3



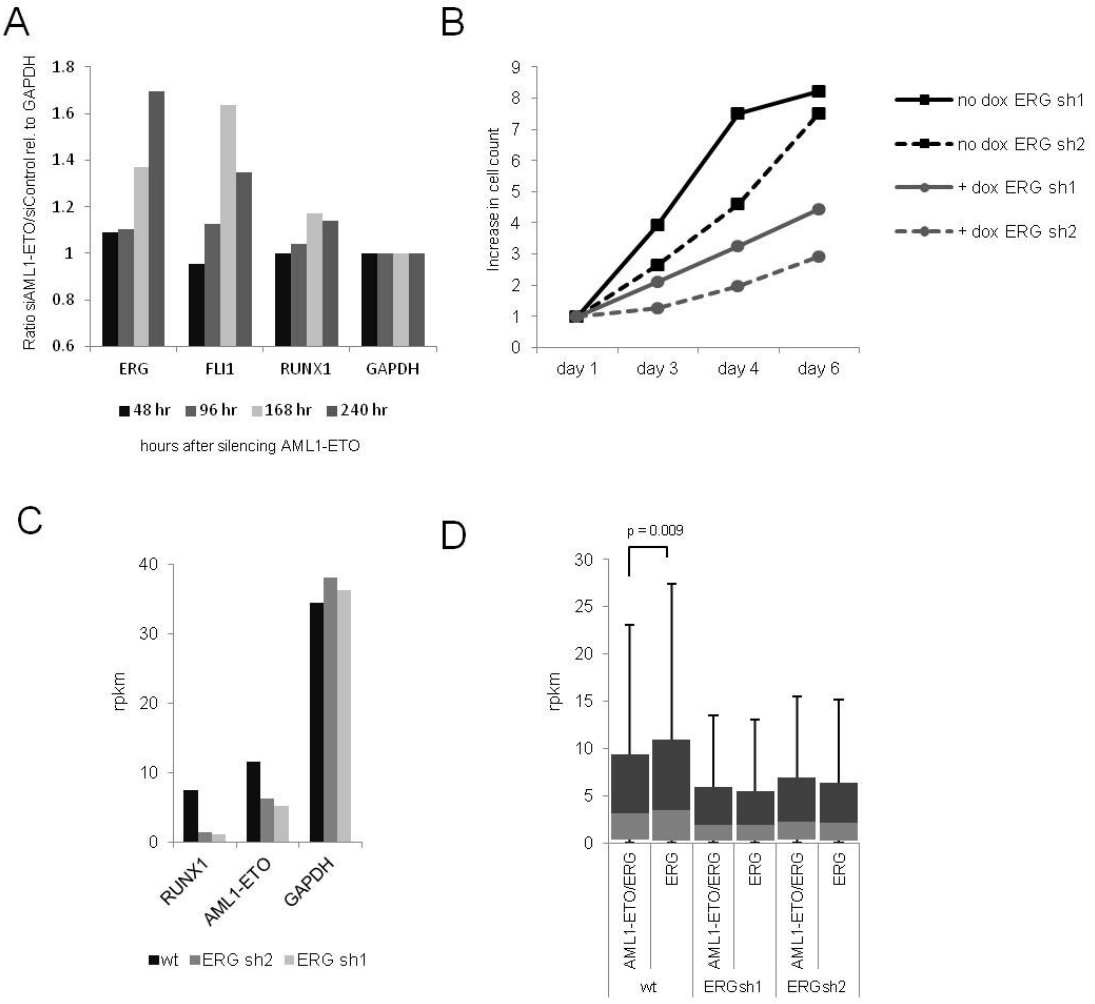
Supplementary Figure S3. A. MCF7 cells transfected with expression constructs for AML1-ETO and ERG or empty vectors. Resultant protein levels were detected using Western blot analysis and antibodies recognizing AML1-ETO, ERG and TBP. B. Transfected AML1-ETO and ERG bind the TYK2 genomic region in MCF7 cells. In blue the AML1-ETO ChIP-seq data is plotted, in yellow the ERG data. C. Venn diagram showing the overlap of AML1-ETO and ERG peaks after MACS peak calling for AML1-ETO and ERG in transfected MCF7 cells. D. UAE cells were treated with zinc to induce AML1-ETO expression. Protein levels were detected using Western blot analysis and antibodies recognizing AML1-ETO and KAP1. E. ChIP-seq using U937 cells expressing (plus zinc) or not expressing (no zinc) AML1-ETO. Overview of the *NFE2* AML1-ETO binding site in U937 AML1-ETO cells. In blue the AML1-ETO ChIP-seq data is plotted and in pink the FLI1 data.

Figure S4



Supplementary Figure S4. A. K562-ERG cells transfected with expression constructs for AML1-ETO or empty vectors and treated or not treated with dox for 72 hours. Resultant protein levels were detected using Western blot analysis and antibodies recognizing AML1-ETO and ERG. B. ChIP-seq using K562-ERG cells expressing high levels (plus dox) or low levels (no dox) ERG and transfected 24 hours before harvesting with AML1-ETO. Overview of the *TAF12* AML1-ETO/ERG binding site in K562-ERG cells. In blue the AML1-ETO (AE) ChIP-seq data is plotted and in yellow the ERG data. C. Boxplot showing the tag density of AML1-ETO and ERG tags in 'increased' ERG binding sites in K562-ERG cells transfected with AML1-ETO and treated or untreated with dox.

Figure S5



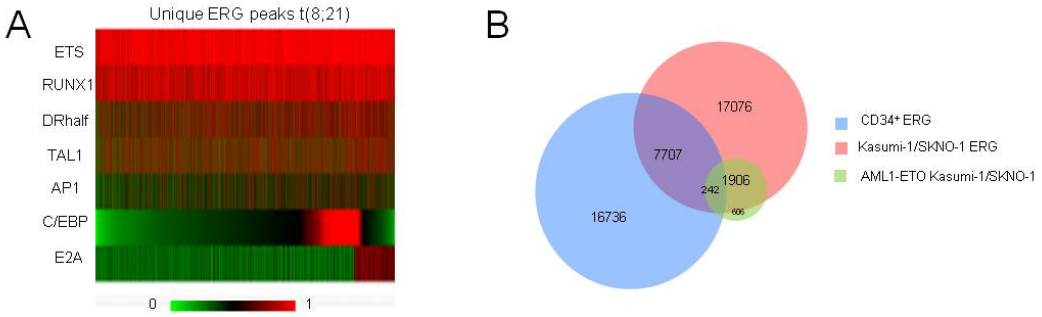
Supplementary Figure S5. A. AML1-ETO silencing leads to altered ERG, FLI1 and RUNX1 expression. Using a previously published expression data set in SKNO-1 cells (Ptasinska et al., 2012), for each time point the difference in expression of the genes was examined between AML1-ETO silenced and control cells and normalized to the difference in expression of the GAPDH gene. B. Growth assay of SKNO-1 ERG sh1 and ERG sh2 cells with and without dox. C. RPKM values of RUNX1 and AML1-ETO in wild-type as well as ERG silenced (+dox) SKNO-1 cells. D. Boxplot representing RPKM values of ERG target genes bound or not bound by AML1-ETO in wild-type SKNO-1 cells as compared to two ERG silenced (+dox) SKNO-1 cell lines.

Figure S6

Enrichment Score: 4.88	**P_Value**
microtubule cytoskeleton	2.5E-7
cytoskeleton	1.2E-6
microtubule organizing center	2.9E-6
centrosome	8.4E-6
cytoskeleton	2.9E-4
Enrichment Score: 4.75	**P_Value**
Ras-association	3.8E-5
Enrichment Score: 4.59	**P_Value**
transcription (co)activator activity	1.6E-6
transcription factor binding	2.8E-6
Enrichment Score: 3.47	**P_Value**
regulation of cell death	7.6E-7
regulation of programmed cell death	1.0E-6
regulation of apoptosis	1.5E-6
Enrichment Score: 3.14	**P_Value**
leukocyte activation	5.2E-6
lymphocyte activation	1.9E-5
leukocyte differentiation	2.6E-4
T cell activation	2.9E-4
lymphocyte differentiation	6.7E-4
B cell activation	2.0E-3
hematopoiesis	3.7E-3
immune system development	6.1E-3

Supplementary Figure S6. Functional annotation clustering (GO) of genes that are associated with AML1-ETO common peaks detected in patient blast cells.

Figure S7



Supplementary Figure S7. A. Heatmap display of enriched motifs present in DNA sequences underlying ERG binding sites that are present in t(8;21) cells but not in $CD34^+$ cells. B. Venn diagram representing the overlap of ERG and AML1-ETO binding sites in Kasumi-1/SKNO-1 cells and ERG binding sites in normal $CD34^+$ cells.

Supplementary Table Legends

Supplementary Table S1, HG18 coordinates
+ page 1. MACS called peak files and target genes of ChIP-seq tracks used in this study.
+ page 2. High confidence AML1-ETO cell line peaks
+ page 3. AML1-ETO target genes affected by ERG silencing
+ page 4. High confidence AML1-ETO peaks from patients

Supplementary Table S2, HG19 coordinates
+ page 1. MACS called peak files and target genes of ChIP-seq tracks used in this study.
+ page 2. High confidence AML1-ETO cell line peaks
+ page 3. AML1-ETO target genes affected by ERG silencing
+ page 4. High confidence AML1-ETO peaks from patients